

Durham Research Online

Deposited in DRO:

22 April 2014

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Jones, K.L. and Tymms, P.B. (2014) 'Ofsted's role in promoting school improvement : the mechanisms of the school inspection system in England.', *Oxford review of education*, 40 (3). pp. 315-330.

Further information on publisher's website:

<http://dx.doi.org/10.1080/03054985.2014.911726>

Publisher's copyright statement:

This is an Accepted Manuscript of an article published by Taylor Francis Group in *Oxford Review of Education* on 25/04/2014, available online at: <http://www.tandfonline.com/10.1080/03054985.2014.911726>.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Ofsted's role in promoting school improvement: The mechanisms of the school inspection system in England

Submitted to: Oxford Review of Education

Authors

Karen L. Jones, Centre for Evaluation and Monitoring (CEM), Durham University
Peter B. Tymms, School of Education, Durham University

Abstract Most countries across Europe now have their own Inspectorate as part of a school improvement and accountability system. However there has been little research on the impact of school inspections or on the aspects of school inspections that maximise the positive effects and minimise the unintended consequences. As a precursor to further research, this paper presents a program theory of the school inspection system in England (Ofsted), highlighting the underlying mechanisms that explain how Ofsted intends to promote school improvement. The program theory is derived in a systematic way using the policy scientific approach, which draws evidence from the multiple methods of interviews and source documents. The key mechanisms that underpin Ofsted's promotion of school improvement were found to be the setting of standards, giving feedback, the use of sanctions and rewards, the collection of information on schools and public accountability. Details showing the logic behind these mechanisms are presented. The program theory is validated by senior personnel from Ofsted. The paper argues for more research in evaluating the impact of school inspection and the specific mechanisms that underpin it.

Key words: inspection, accountability, program theory, policy scientific approach

Introduction

The majority of countries across Europe, have implemented systems of evaluation and accountability in order to ensure quality of education in schools (Eurydice, 2004). These systems (which also include for example testing and the publication of league tables) primarily take the form of school inspections, with most countries having their own Inspectorate. Despite this, and given the vast sums of public money spent on school inspection (in England alone in 2011/12 £167 million was spent on the inspection and regulation of education, children's services and skills (Ofsted, 2012)) there is surprisingly little firm evidence for the impact of school inspections in the literature and on how school inspection is supposed to promote school improvement. Most of the research has been in England and the Netherlands but findings have been mixed (see for example De Wolf and Janssens, 2007; Shaw *et al*, 2003; Rosenthal, 2004; Luginbuhl *et al*, 2009).

The systems employed by inspectorates to drive school improvement vary across Europe, ranging from systems involving sanctions with governmental control to systems with few consequences based on peer review, and from emancipatory systems involving self-evaluation to bureaucratic systems involving compliance and regulation (Faubert, 2009; Hughes *et al*, 1997; McGarvey and Stoker, 1999; in Ehren *et al*, 2012).

The European Commission, Lifelong Learning Project, 'The impact of school inspection on teaching and learning', of which this study is part, aims to fill some of the gaps in this knowledge, by comparing differing inspectorates in six European countries in order to evaluate the impact of inspections and to attempt to identify the mechanisms that maximise a positive impact, and minimise any unintended effects. This comparative research started with a clear description of the conceptual model of the inspectorate in each country. Whether one is a proponent of theory-driven evaluation, with the aim of testing theoretical models with empirical data (e.g. Weiss, 1997; Chen, 1990), or prefers theory-driven evaluation alongside randomised experiments designed to evaluate impact mechanisms (Cook, 2000), few can disagree with the value of developing a program theory (or conceptual model) of a program that is going to be evaluated. This is the purpose of this paper, which develops a conceptual model describing the mechanisms and assumptions that underpin the inspection system of state schools in England: The Office for Standards in Education, Children's Services and Skills (Ofsted). Ofsted has a number of roles, including providing accountability, ensuring compliance with national regulations, informing consumer choice, providing value for money and promoting service (including school) improvement (Ofsted, 2005; England, 2006). This paper will focus on school improvement, and the resulting conceptual model will set out how Ofsted aims to promote school improvement. Such a model has not been presented in literature before, however Ofsted (2005) presents a number of ways that inspection can have an impact. The model here will be developed in a systematic way using the 'policy scientific approach' (Leeuw, 2003). This is the first time an explicit conceptual model for Ofsted is set out in the literature and it has the advantage of having been validated by senior personnel at Ofsted.

The policy scientific approach enables us to reconstruct the mechanisms that explain how the school inspection system in England aims to promote school improvement. The collection of mechanisms forms a 'program theory' (Chen, 1990 in Ehren *et al*, 2005) that explains how inspection is expected to bring about school improvement. The program theory is then critically analysed to predict the (in)effectiveness of inspections in promoting improvement. This paper focuses on inspections in maintained schools in England in the age range of 4-16 years (Reception through to Year 11). The program theory described here necessarily focusses on one particular period in history and is based on the framework for school inspection that ran from September 2009 until August 2011. Ofsted is constantly evolving and has already moved on, although it can be argued that much of its essence is fairly constant.

The paper aims to answer the following three research questions:

- What are the mechanisms by which Ofsted's actions are expected to improve the education of children and young people?
- How consistent, precise and realistic are the mechanisms regarding the expected effects of inspections on promoting school improvement?
- How likely are inspections to successfully promote school improvement?

The paper will begin with a brief overview of the method used to analyse the inspection system (program theory and policy scientific approach). The results of the various stages of the policy scientific approach will then be presented, ending with a program theory based on the assumptions underlying the government's aim to

promote school improvement. The program theory will then be analysed in order to predict its (in)effectiveness. The paper concludes with a summary/discussion.

Method

A number of methods exist for reconstructing the theories and assumptions that underpin a program (see Leeuw, 2003). The method selected for this paper is the policy scientific approach (as outlined in Leeuw, 2003 and Ehren *et al*, 2005; and similar to the method described by Renger, 2011). This method is empirical and analytical in nature drawing strength from its reliance on multiple data sources (documents and interviews) as well as its use of diagrams to present the program theory (Leeuw, 2003). The approach consists of a number of steps which are presented below:

Step 1: Reconstruction of the Program Theory

This step involves reconstructing the mechanisms that explain how a program or policy aims to bring about its goals (in this case the mechanisms that explain how inspection aims to promote school improvement). The reconstructed mechanisms in turn form the reconstructed program theory. The reconstruction of the program theory is split into three stages:

- i) The aims of the program theory are stated.
- ii) Information is gathered from which the program theory will be constructed.
- iii) Information is translated from the source documents into mechanisms that will form the program theory. This involves searching for statements in the source documents that link the mechanisms to the goals of the program. These statements are then reformulated into 'if-then' statements and the collection of these reformulated statements/mechanisms forms the program theory.

Step 2: Validation of the program theory

Once the program theory is reconstructed it needs to be validated. In this instance the program theory was validated with relevant senior personnel from Ofsted to check whether the program theory is accurate.

Step 3: Evaluation of the program theory

The final step in the method is to evaluate the program theory. The program theory will be analysed for consistency, completeness and realism in order to estimate the mechanisms' potential for meeting the intended effects of promoting school improvement. There will be a check of the consistencies in the logic of the mechanisms and a check of whether the mechanisms are clear or have any gaps. The mechanisms are also analysed for realism using evidence from literature.

Results

The aim of the program theory presented in this paper is to describe and evaluate the assumptions that underpin Ofsted's role in promoting school improvement for the period September 2009 until August 2011. The program theory is based on two levels of information/evidence: explicit information in the Ofsted literature and Acts of Parliament, and implicit mechanisms commonly believed by Ofsted as an organisation but not explicitly written down. An insight into these mechanisms can be gained through comments made by staff at Ofsted through interviews as part of the validation process and through speeches made by the Chief Executive of Ofsted (Christine Gilbert) at the time. The evidence at this level varies in that different members of staff may have different opinions as to the role Ofsted plays in driving improvement.

Source documents

The key functions of Ofsted are set out in law through two key Acts of Parliament – the Education Act 2005 (England, 2005) and the Education and Inspection Act 2006 (England, 2006) (elements of which supersede the 2005 Act). In addition to the requirements of law, details on the frameworks and mechanisms used by Ofsted in order to meet its aims are set out in key Ofsted documents. The conceptual model presented in this paper is based on evidence from the following documents:

1. Education Act 2005 (England, 2005)
2. Education and Inspections Act 2006 (England, 2006)
3. The Framework for School Inspection (Ofsted, 2009a)
4. Ofsted Inspects (Ofsted, 2009b)
5. Ofsted, Raising standards, improving lives (Ofsted, 2009c)
6. The Evaluation schedule for schools (Ofsted, 2010)

Overview of Ofsted as defined by Education Acts

Prior to the Education and Inspections Act 2006 the primary function of Ofsted as outlined in the 2005 Education Act was to keep the Secretary of State informed about matters connected to the quality of Education provided by schools in England (and Wales in a separate section of the Act). Examples of such matters include educational standards achieved in schools, the quality of the leadership and management of schools and the behaviour and attendance of pupils. It was set in law that the Chief Inspector should make an annual report, and provide additional reports where needed, to the Secretary of State on the state of English schools. The Act gave Ofsted the right of entry into schools, including access to pupils and records. It was also stated that all schools in England (excluding private (fee-paying) schools) should be inspected within a given time period. Regulations from September 2009 state that all maintained schools must be inspected every 5 years. These compulsory inspections are known as 'Section 5' inspections as they are specified in Section 5 of the Education Act.

In addition to Section 5 inspections, Section 8 of the Act gave the inspectorate the power to perform further inspections at the discretion of the inspector. These inspections are known as 'Section 8' inspections.

The Act also specified that where schools were found lacking (i.e. schools causing concern) in the Section 5 inspections they should be categorised as being in either 'special measures' if the school is failing and does not have the capacity to improve (as judged by Ofsted) or given a 'notice to improve' where schools are required to improve as they are performing 'less well than expected' (Section 13, England, 2005). Schools in special measures are required to work with the local authority (Section 15, *ibid*) and if no improvement is seen they are under threat of closure by the Secretary of State (Section 45, *ibid*).

The 2005 Act as it stood did not state any direct functions related to the improvement of schools or the raising of standards in schools. Apart from the power to put a small proportion of failing schools into special measures and demand improvement, the exercise of inspecting *all* schools in England was explicitly for the purpose of informing government.

The majority of the 2005 Education Act that refers to inspection still held for the period covered by this paper and still holds at this time of writing, however in 2006 the Education and Inspections Act rewrote the main functions of Ofsted and superseded the functions stated in the 2005 Act. The new functions of Ofsted are set out in Sections 116 to 119 of the Act. Sections 116 and 118 show that the key functions of Ofsted are still to keep the Secretary of State informed about the quality of schools and to give advice. For example:

The Chief Inspector has the general duty of keeping the Secretary of State informed about—

- (a) the quality of activities within the Chief Inspector's remit and (where appropriate) the standards achieved by those for whose benefit such activities are carried on,*

Section 118 (1) (a), England (2006)

However the 2006 Act places a new emphasis on promoting improvement (as well as ensuring services are user-focussed and provide value for money):

(1) The Office is to perform its functions for the general purpose of encouraging—

- (a) the improvement of activities within the Chief Inspector's remit,*

Section 117 (1) (a), England (2006)

Clearly, one of Ofsted's overarching aims is to promote improvement, however there is not a tightly coupled causal link which states that Ofsted aims to improve standards in schools, rather it aims to 'encourage' (i.e. promote) improved standards. However, as a result of promoting improvement, we would expect to see improvement in a proportion of schools. The mechanisms of school inspections that Ofsted aim to use to promote improvement are presented in the next section.

Background information on Ofsted gradings and monitoring inspections from 'The evaluation schedule for schools' (Ofsted, 2010)

Specific information on the gradings awarded to schools following inspections are not found in the Acts, but can be found, along with grade descriptors, in the Ofsted document ‘The evaluation schedule for schools’ (Ofsted, 2010). As a result of Section 5 inspections schools are graded as either ‘outstanding’, ‘good’, ‘satisfactory’ or ‘inadequate’. Schools graded as inadequate are further split into schools with a notice to improve and schools in ‘special measures’. All schools graded as inadequate and a proportion of schools graded as satisfactory receive ‘Section 8’ monitoring inspections. The timings and frequency of Section 8 inspections vary according to whether the school is satisfactory, has a notice to improve or is in special measures. There are therefore three thresholds for three different monitoring processes, numbered 1, 2 and 3 on Figure 1 below. Figure 1 also presents the proportion of schools and the number of schools placed in each category in 2008/09.

Figure 1: Summary of Ofsted inspection gradings and thresholds, annotated with figures and proportions from 2008/09 Section 5 inspections

19% ~1011 primary schools 203 secondary schools Outstanding	50% ~2661 primary schools 535 secondary schools Good	28% ~1490 primary schools 300 secondary schools Satisfactory	4% ~213 primary schools 43 secondary schools Inadequate	
		~40% of schools judged as satisfactory were monitored	Notice to improve	Special measures
		1	2	3

Information on the standards against which schools are judged during inspections

‘The framework for school inspection’ (Ofsted, 2009a) sets out an overview of the standards that are set and specific details are presented in the ‘Evaluation Schedule for schools’ (Ofsted, 2010). In summary, schools are judged and graded on seven key areas related to pupil outcomes (including attainment, progress and behaviour), three judgements concerning the effectiveness of provision (for example the quality of teaching) and eight judgements related to the leadership and management of schools.

Evidence linking the mechanisms of Ofsted to the goal of promoting school improvement

Documentation evidence

As noted above Section 117 of the ‘Education and Inspection Act 2006’ (England, 2006) makes it clear that one of the primary aims of Ofsted is to encourage the improvement of schools, and the education system as a whole through inspection. The document ‘The Framework for School Inspection’ (Ofsted, 2009a) explains the mechanisms by which Ofsted aims to use school inspections to promote the improvement of schools. Further mechanisms are highlighted in ‘Ofsted, Raising

standards, improving lives' (Ofsted, 2009c). Details on how Ofsted aims to maximise the effect of promoting improvement through inspections are found in the document 'Ofsted Inspects' (Ofsted, 2009b).

'The Framework for School Inspection' (Ofsted, 2009a) makes it explicit that school inspections aim to promote improvement in four key ways. Firstly, by setting expectations, schools are given a description of the standards they are expected to meet, and as highlighted in Ofsted (2005), the anticipation of an inspection provides an impetus to meet these standards. Secondly, by increasing a school's confidence by endorsing the school's own view of its effectiveness (if it is accurate), or by providing a 'sharp challenge (and an impetus to act)' if improvement is needed. Thirdly, by recommending priorities for school action, and in some cases following up progress, although emphasis is placed on the capacity of school to actually respond to these recommendations (*ibid*). Finally inspection is considered to promote improvement through the conversations between inspectors and senior managers at school during the inspection. The inspection process is considered to complement the school's self-evaluation and promote its rigour.

Explicit statements that link the placing of schools in special measures/notice to improve/satisfactory with extra monitoring to school improvement are difficult to find, although it is implicit that local authority involvement, extra monitoring etc are intended to promote improvement.

'Ofsted, raising standards, improving lives' (Ofsted, 2009c) makes it clear that as well as influencing schools, inspections can promote improvement by influencing policy-thinking and policy-making. Evidence collected through inspections is disseminated via survey reports on a range of topics, including subject reports and reports of other aspects of learning in schools. As well as reports, conferences and other publications are provided for schools in order to highlight areas of good practice, e.g. 'Twelve Outstanding Secondary Schools' (Ofsted, 2009d).

Interview evidence

Although difficult to find explicitly in key documents, it was clear from our initial interview meeting with key personnel at Ofsted that an important mechanism for driving improvement is related to the publication of inspection reports. Any parent can view a report and they can use this information to inform their decisions about what school to apply to for their child. This adds market forces into the education system. Parents can choose the good schools and avoid the weak schools. The argument is that by making the reports public schools will do what they can to improve on standards, not only to save face, but also because school admissions can have a big effect on job security. As school funding is linked to the number of pupils, parents 'flocking' to a particular school can give job security to school staff, whereas a poor report that leads to parents avoiding a school can lead to job insecurity and job losses. Schools should be highly motivated to avoid being on the wrong side of inspection thresholds.

The validation meeting highlighted powerful incentives for schools to be judged as 'outstanding'. These incentives, in effect, act as rewards. They include the fact that outstanding schools can apply for 'Academy' status, they can become 'teaching schools' and head teachers from outstanding schools can apply to become national leaders of education (NLEs). There is some evidence that these incentives are driving

change as Ofsted is being contacted by non-outstanding schools asking to be re-inspected.

The validation process also highlighted the fact that inspections could be triggered sooner if, for example, serious complaints were made about a school. For a small number of schools, it was thought that this threat may be enough to keep standards from slipping.

Further information on how improvement occurs was highlighted in the interview/validation meeting. It was made clear that school improvement does not happen on its own – it happens in relationship with other drivers of improvement. Ofsted's improvement role is dependent on other actors, with the three key actors being parents, policy makers and schools.

The role of parents is seen to be the primary determinant of pupil outcomes, including attainment. A school's engagement with parents is therefore critical. The parents' role in driving improvement surrounds public availability of inspection reports. There are two mechanisms involved here. Firstly, parents can use the reports when deciding which school to apply to for their child. This introduces market forces. Parents use the reports to 'vote with their feet' so that strong schools grow and become oversubscribed and weak schools receive fewer applicants. This mechanism (the publication of reports) is seen by schools as being more of a sanction/reward than the threat of school closure. School closure resulting in job losses affects a small minority of schools whereas admissions has more of an immediate impact on the reputation of the school and changes in admissions can affect the security of jobs. A second mechanism is that parents may pressurise schools directly to improve.

The mechanisms that directly involve schools operate at two levels (head teacher/senior management team level and the teachers themselves). A lot of emphasis is placed on the relationship between the head teacher and the inspectors. The head teacher will often accompany the inspectors as they observe lessons. A dialogue will occur between the head teacher and the inspector in order to gauge whether the two of them make the same judgements as each other. This can amount to mentoring the head teacher. A lot of emphasis is placed on ascertaining whether the head teacher and senior management have an accurate view of the school, its strong points and its areas for improvement. If a school has an accurate view of itself then it is more likely to be able to identify areas for improvement and put into practice plans for development. At the teacher level, a sample of lessons is observed during an inspection and feedback is available to the teachers who are observed.

Ofsted is reliant on these actors. They cannot directly make schools do anything apart from the small minority of schools that may close as a result of being in special measures and not improving. Inspection is a 'nudge' (Thaler and Sunstein, 2008) mechanism. Similarly, peer review may also act as a mechanism involving actors other than Ofsted.

Reformulation of the statements into conditional 'if-then' type propositions

Statements from documents and interviews have been reformulated into conditional 'if-then' statements to describe the mechanisms that lie behind the goal of promoting school improvement. These statements have been grouped into five broad mechanisms and are presented in Table 1.

Table 1: The assumptions underlying school inspections in England

Setting Standards

1. If schools are given details (criteria and descriptors) on the standard of performance expected of schools then they will attempt to meet those expectations and this will have a positive impact on average.
 - 1.1 If standards are set, then schools can complete the self-evaluation form (SEF) in order to rate themselves against standards.
 - 1.2 If accurate, this form will allow schools to identify areas for improvement.
 - 1.3 If areas are identified for improvement then a development plan can be created.
 - 1.4 If schools have a development plan then the development plan can be put into action.
 - 1.5 If the plan is put into action effectively then schools will improve as a result.

Giving Feedback

2. If schools are given feedback about strengths and weaknesses they will use this feedback to improve.
 - 2.1 If inspections endorse the schools own view of their effectiveness (when it is accurate) it will lead to improved confidence which will in turn lead to improvement on average. If a school does not judge itself accurately then a sharp challenge is given and an impetus act which will lead to improvement.
 - 2.2. If Ofsted recommends priorities for future action and checks progress where it is deemed appropriate schools will generally improve.
 - 2.3 If a constructive dialogue between inspectors and the senior leaders and staff of the school is established the school will be more likely to improve.
 - 2.4 If precise actions to underpin the recommendations are identified then schools are more likely to improve.

Sanctions and Rewards

3. If schools receive sanctions and rewards they will be motivated to improve
 - 3.1 If (a small minority of) failing schools are placed in special measures then the schools will improve in order to move out of special measure and to avoid sanctions (e.g. job losses and school closure)
 - 3.2 If outstanding schools are given rewards then schools will improve to become outstanding in order to receive those rewards

Collecting information

4. Collecting information
 - 4.1 If information gathered from inspections is passed on to policy makers then it will inform their thinking and decision making which will lead to policies that promote improvement.
 - 4.2 If examples of good practice observed during inspections are disseminated through reports, conferences and other publications this will promote school improvement within the schools themselves.

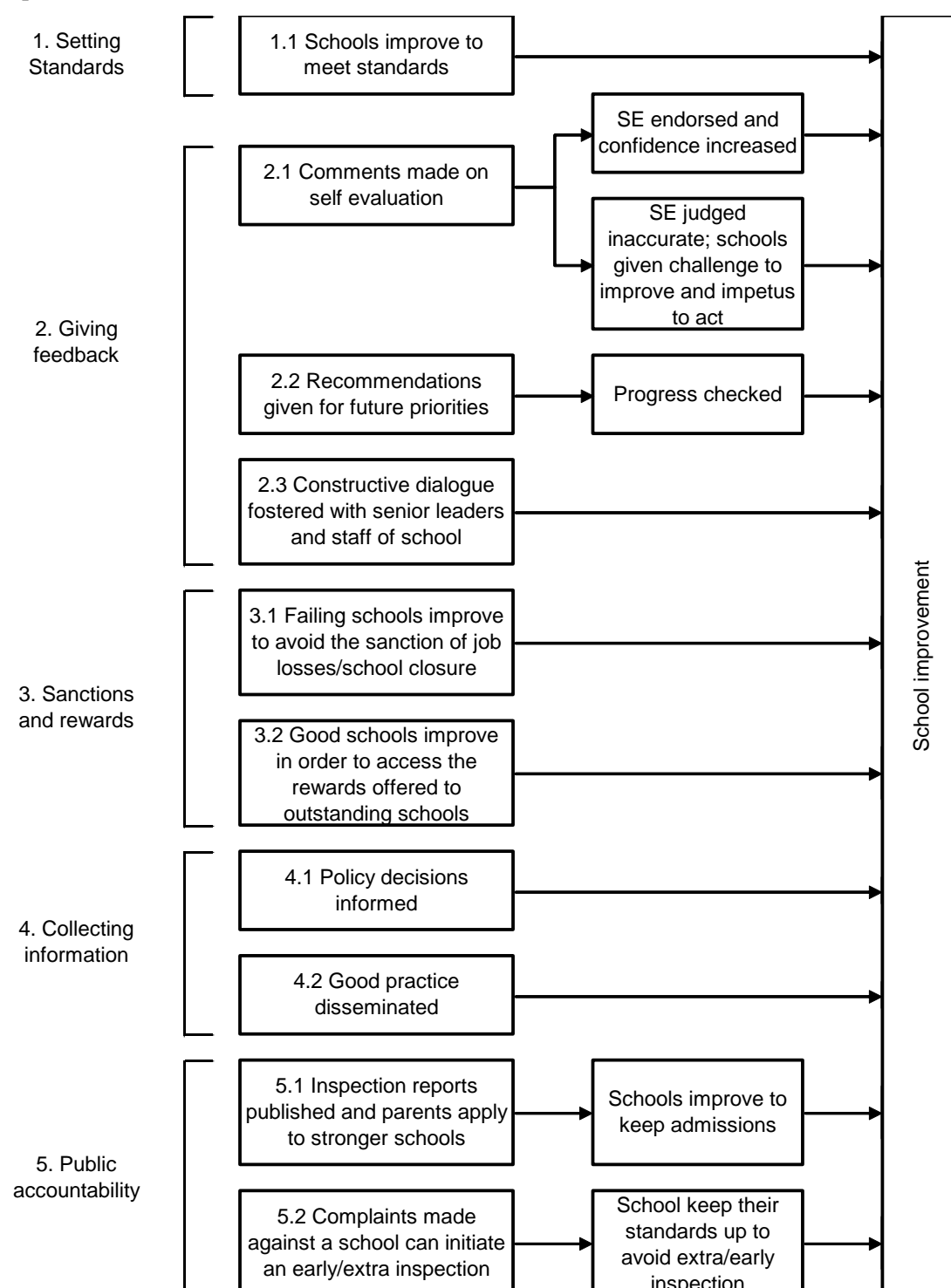
Public Accountability

5. If inspection reports are published then schools will be motivated to improve in order to save face and remain /become popular.
 - 5.1 If reports are published parents will read them or come to know of their contents
 - 5.2 If parents know about the content of reports they can make informed choices about what schools to apply to for their child
 - 5.3 If a school gets a good Ofsted report more parents will want to apply to that school, and conversely, if school receives a poor Ofsted grading parents will want to avoid that school
 - 5.4 If school applications increase then a school will become oversubscribed and there will be job security for school staff. If school applications decrease then a school will become undersubscribed and as school funding is based on pupil numbers jobs will become insecure and/or there will be job losses.
 - 5.5 If the publishing of inspection reports affects job security then schools will improve in order to get good inspection results and enhance job security.
 - 5.6 If it is known that inspections can be initiated at any time due to 'whistleblowers' then schools will keep their standards up

Summary of the Program Theory

The mechanisms described in Table 1 can be placed together graphically to present a program theory that demonstrates how inspection intends to promote improvement – see Figure 2. It is important to highlight here again that the program theory relies on the role of actors, other than the inspectorate, such as parents, policy makers and schools.

Figure 2: Program theory to show how Ofsted inspections are expected to promote school improvement



Validation of the Program Theory

A face-to-face validation meeting with personnel from Ofsted took place at Ofsted Headquarters in London on 17th March 2011. A follow-up telephone conference call took place on 21st March 2011.

At the face-to-face meeting the first draft of the program theory was presented and discussed in order to check whether it is was an accurate representation of the mechanisms that are intended to promote school improvement. The program theory was adjusted and presented back to Ofsted staff prior to the telephone conference call. The program theory presented above is the final version after consultation.

Evaluation of the Program Theory

The final stage of the policy scientific approach is to evaluate the program theory. This involves critically evaluating the steps in the logic in order to identify any gaps, i.e. are the mechanisms complete and consistent? It also involves evaluating the program theory for realism. This involves searching the literature for evidence that indicates whether or not the program theory will be successful.

Consistency

Here the question is asked as to whether there are any logical inconsistencies in the mechanisms of the program theory. Inspection in England is considered to be a 'nudge' mechanism (Thaler and Sunstein, 2008), meaning it does not directly cause improvement, but rather nudges other actors into improving schools. This corresponds to the stated position that Ofsted 'promotes' school improvement. It relies on schools, parents and policy makers to do the improving. Schools do not have to act on feedback and advice given during inspection (apart from the small minority of schools in special measures). It is however assumed that some schools will act on feedback and suggestions for improvement.

One possible inconsistency lies with the mechanisms suggesting that market forces in education lead to school improvement. This could be a 'zero sum game' – some schools get better whilst others decline – but it is acknowledged that the view is that more would improve than are harmed.

Completeness

Here the mechanisms underlying the program theory are assessed to investigate whether they are clearly defined or leave gaps. The findings are that in general the mechanisms are clearly defined.

Realism

Searching the literature for each of the mechanisms in the program theory is a vast undertaking, not within the scope of this paper. The literature alone on feedback is

voluminous. Many have proposed models on feedback and suggested the conditions under which feedback is maximised (see for example Hattie and Timperley, 2007; Coe, 1998; Black and Wiliam, 1998; Brimblecombe *et al*, 1996; Brinko, 1993). What we do know is that the mechanisms are complex and do not show simple solutions. Coe (1998) states that theories based on good empirical evidence that link the effects of feedback on performance are limited and care needs to be taken in translating these inferences from one context to another. He argues that the view that *'those who receive feedback on a task perform better than those who do not'* cannot be universally accepted from the literature. The literature does not give a simple answer as to whether feedback will lead to school improvement. Hattie and Timperley (2007) provide a detailed theoretical model which aims to give a clear account of when feedback is likely to have a positive effect so far as student learning is concerned. Their account makes it clear that the possibility of feedback having negative outcomes is very real. Unfortunately we lack a similar account, and the empirical data, for feedback from inspections.

Even when the literature is vast, the findings can be inconclusive and not necessarily specific to the school level. But there are also aspects of the program theory where there is no empirical evidence in the literature to illustrate if the particular mechanism works or not. An example is the key mechanism that states that inspectors' endorse or challenge the school's self-evaluation in order to motivate and encourage school improvement.

Clearly more research is needed, which focusses on the individual mechanisms of inspection, aimed at the school level. In light of this, the best that can be done here is treat the complete program theory as a black box and ask what impact inspection has had. A number of studies have attempted to measure the effects of inspection visits on pupil attainment, arguably a key indicator of school quality (Gray and Wilcox, 1995). The studies have mainly been carried out in England and the Netherlands and the findings have been mixed. An Ofsted report (Matthews and Sammons, 2004), looking at schools inspected over the period of 1993-2002, claims that, on average, school performance in national exams (Key Stage 4) at age sixteen is better in the two years following inspection (paragraph 68). In contrast, Gray and Wilcox (1995) and Cullingford and Daniels (1999) found that Ofsted inspections have a negative impact on the proportion of pupils achieving 5A*-C grades at GCSE in England, although prior attainment is not taken into account in their design, external validity is weak and the authors themselves point out that they cannot claim causality. Shaw *et al* (2003) and Rosenthal (2004) used more sophisticated methods and again, both conclude that Ofsted inspection had a small negative impact on GCSE results (taken at age 16) in state secondary schools in the year following inspection. Rosenthal postulates that this is due to schools focussing on Ofsted standards during the year of inspection, perhaps at the expense of student attainment. As De Wolf and Janssens (2007) point out, these studies do not look at the longer term impact of inspection or the impact in other sectors of education, including primary schools. In addition, they all relate to earlier versions of Ofsted inspections. More recently, using more sophisticated and stronger designs, Hussain (2012) and Allen and Burgess (2012) have focussed on the impact of inspections on the minority of schools that fail inspections (i.e. those that receive an 'unsatisfactory' grading in their main Ofsted inspections). Hussain focusses on the national Key Stage 2 tests taken at the end of Primary School (age 11) and found that failing an inspection led to increased test gains of around 0.1 of a standard deviation.

Allen and Burgess found that schools that just fail their inspection see an improvement in test scores two and three years after inspection. The positive impact was mainly found of middle and higher ability pupils. Outside of England, Luginbuhl *et al* (2009) studied the effects of Dutch primary school inspections on test scores finding small positive effects and no effects using two different designs.

As documented in De Wolf *et al* (2007) it is widely accepted that school inspections and accountability systems in general have side-effects/unintended effects (see for example Smith, 1995; Fitz-Gibbon, 1997; Chapman, 2001; Koretz, 2002). One key example of unintended strategic behaviour is 'myopia', described by Smith (1995) as the focus on short-term solutions at the expense of long-term improvement, for example schools changing their curriculum away from academic subjects that are perceived as difficult towards vocational subjects that are perceived as easy in an attempt to increase examination results. Another key unintended effect is 'teaching to the test' or 'indicator fixation' where teachers have excessive concentration on exam performance (see Fitz-Gibbon (1997) and Koretz (2002)). Leeuw (2000) suggests that the undesired effects may have the capacity to undo any of the positive effects of school inspections. In light of this, any study of the impact of school inspections needs to take into account the impact of any undesired effects.

Discussion/Summary

This paper presents a clear, validated model of the assumptions the lie behind Ofsted's aim to promote school improvement. This provides a structure within which the impact of inspection can be assessed. At the moment there is a lack of evidence from strong research designs to assess the impact of inspections and the assumption that there is a causal link between inspections and school improvement cannot be clearly supported from the literature. Research is needed, using strong research designs aimed at investigating the causal link between inspections and school improvement for all schools, including measuring both the short term and long term impact of inspections. Research that provides more evidence on the potential success of the individual mechanisms in the program theory would be particularly valuable. This research would need to take into account unintended consequences of school inspection. This research could be used to improve the existing system, but also to inform changes to the structure of the system itself.

Notes on contributors

Karen Jones is a Research Associate in the Centre for Evaluation and Monitoring (CEM) at Durham University. Her research interests include school accountability, assessment and quantitative research methodology.

Peter Tymms is Head of Department in the School of Education at Durham University. His main research interests include monitoring, assessment, interventions and research methodology generally. He set up the PIPS (Performance Indicators in Primary Schools) Project which runs in thousands of schools around the world.

References

- Allen, R. & Burgess, S. (2012) *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England*, CMPO working paper No. 12/87.
- Black, P. J. & Wiliam, D. (1998) *Inside the black box: raising standards through classroom assessment* (London, King's College London School of Education).
- Brimblecombe, N., Shaw, M. & Ormston, M. (1996) Teachers' intention to change practice as a result of Ofsted school inspections, *Education Management and Administration*, 24, 339-354.
- Brinko, K. T. (1993) The practice of giving feedback to improve teaching: what is effective? *Journal of Higher Education* 64(5): 574-593.
- Chen, H. (1990) *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chapman, C. (2001) Changing classrooms through inspections, *School Leadership and Management*. 21, 59-73.
- Coe, R. (1998). Can Feedback Improve Teaching? *Research Papers in Education* 13(1): 43-66.
- Cook, T. D. (2000) The false choice between theory-based evaluation and experimentation. In P.J. Rogers, T. A. Hacsí, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities. New directions in evaluation* (Vol. 87, pp. 27-35). San Francisco: Jossey-Bass.
- Cullingford, C. L., & Daniels, S. (1999) Effects of Ofsted inspections on school performance. In C. Cullingworth (Ed.), *An inspector calls: Ofsted and its effect on school standards* (pp. 59-69). London: Kogan Page.
- De Wolf, I. F. & Janssens, J. G. (2007) Effects and side effects of inspections and accountability in education: an overview of empirical studies, *Oxford Review of Education* Vol. 33, No. 3, July 2007, pp. 379-396.
- England (2006) *Education and Inspection Act (2006): Chapter 40* [online] London : HMSO.
- England (2005) *Education Act (2005): Chapter 18* [online] London : HMSO.
- Ehren, M. C. M., Altrichter, H., McNamara, G. & O'Hara, J. (2013) Impact of school inspections on improvement of schools – describing assumptions on causal mechanisms in six European countries, *Educational Assessment, Evaluation and Accountability*, 25, 3-43.
- Ehren, M. C. M., Leeuw, F. L., & Scheerens, J. (2005) On the impact of the Dutch educational supervision act; Analyzing assumptions concerning the inspection of primary education, *American Journal of Evaluation*, 26(1), 60-76.
- Eurydice (2004). Evaluation of Schools providing Compulsory Education in Europe. <http://www.eurydice.org/portal/page/portal/Eurydice>.
- Faubert, V. (2009) School evaluation: current practices in OECD countries and a literature review. OECD Education Working Papers, No. 42, OECD Publishing. doi:10.1787/218816547156.

- Fitz-Gibbon, C. T. (1997) *Feasibility studies for a national system of Value Added indicators* (London, SCAA).
- Gray, J. & Wilcox, B. (1995) *Good school, bad school*. Milton Keynes: Open University Press.
- Hattie, J. & Timperley, H. (2007) The Power of Feedback, *Review of Educational Research* 77(1): 81-112.
- Hughes, G., Mears, R., & Winch, C. (1997) An inspector calls? Regulation and accountability in three public services. *Policy and Politics*, 25(3), 299–314.
- Hussain, I. (2012) *Subjective performance evaluation in the public sector: Evidence from school inspections*, CEP discussion paper CEEDP0135.
- Koretz, D. (2002) Limitations in the use of achievement tests as measures of educators' productivity. In E. Hanushek, J. Heckman, and D. Neal (Eds.), *Designing Incentives to Promote Human Capital*. Special issue of *The Journal of Human Resources*, 37 (4), Fall.
- Leeuw, F. L. (2000) Onbedoelde neveneffecten van outputsturing, controle en toezicht?, in: Raad voor Maatschappelijke Ontwikkeling, *Aansprekend burgerschap; de relatie tussen de organisatie van het publieke domein en de verantwoordelijkheid van de burgers* (Den Haag, Raad voor Maatschappelijke Ontwikkeling), 151–171. [Unintentional side effects of output manipulation, control and school inspections?, in: *Responsible citizenship; the relationship between the organisation of the public domain and the responsibility of the citizens*].
- Leeuw, F. L. (2003) Reconstructing program theories: methods available and problems to be solved, *American Journal of Evaluation*, 24(1), 5–20.
- Luginbuhl, R., Webbink, D. and de Wolf, I. (2009) Do inspections improve primary school performance? *Educational Evaluation and Policy Analysis*. September, 2009, Vol. 31, No. 3, pp. 221-237.
- Matthews, P., & Sammons, P. (2004). *Improvement through inspection: An evaluation of the impact of Ofsted's work*. London: Ofsted (HMI 2244).
- McGarvey, N., & Stoker, G. (1999) *Intervention, inspection, regulation and accountability in local government*. DETR – Interim literature review. London: DETR.
- Ofsted (2005) Office for Standards in Education. *The Annual Report of Her Majesty's Chief inspector of Schools 2004/05*. London
- Ofsted (2009a) Office for Standards in Education. *The framework for school inspection: The framework for school inspection in England under section 5 of the Education Act 2005, from September 2009*. London.
- Ofsted (2009b) Office for Standards in Education. *Ofsted inspects: A framework for all Ofsted inspection and regulation*. London.
- Ofsted (2009c) Office for Standards in Education. *Ofsted, raising standards, improving lives*. London.
- Ofsted (2009d) Office for Standards in Education. *Twelve outstanding secondary schools: Excelling against the odds*. London.
- Ofsted (2010) Office for Standards in Education. *Evaluation schedule for schools*. London.

- Ofsted (2012) *Office for Standards in Education, Children's Services and Skills: Annual Report and Accounts 2011-12*. London.
- Renger, R. (2011) Constructing and verifying program theory using source documentation, *The Canadian Journal of Program Evaluation*, 25 (1), 51-67.
- Rosenthal, L. (2004) Do school inspections improve school quality? Ofsted inspections and school examination results in the UK, *Economics of Education Review*, 23, 143-151.
- Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003) Do Ofsted inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal*, 29, 63-75.
- Smith, P. (1995) On the unintended consequences of publishing performance data in the public sector, *International Journal of Public Administration*, 18, 277-310.
- Thaler, R. H. & Sunstein, C. R. (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness* (New Haven, CT: Yale University Press).
- Weiss, C. (1997) 'Theory-based Evaluation: Past, Present and Future', in D. J. Rog (ed.) *Progress and Future Directions in Evaluation, New Directions for Evaluation*, 76-76. San Francisco, CA: Jossey-Bass.